R E P O R T R E S U M E S

ED 013 538 56 EM 000 488

THE RELATIVE EFFECTIVENESS OF SUPPLEMENTING PROGRAMED INSTRUCTION WITH BLOCKED VERSUS SPACED REVIEW.

BY- SCANLON, JAMES A. TOM, FREDERICK K.T.
CORNELL UNIV., ITHACA, N.Y.

REPORT NUMBER BR-7-8068 PUB DATE 18 MAY 67
GRANT OEG-1-7-078068-2993
EDRS PRICE MF-$0.25 HC-$1.80 45P.

DESCRIPTORS- *PROGRAMED INSTRUCTION, *REVIEW (REEXAMINATION), *GRADE 9, *AGRICULTURAL EDUCATION, *INSTRUCTIONAL FILMS, PROJECTION EQUIPMENT, PROGRAMED TEXTS, LEARNING, RETENTION, EXPERIMENTS

NINTH GRADE STUDENTS ENROLLED IN FIRST-YEAR VOCATIONAL AGRICULTURE IN TWENTY-SEVEN NEW YORK SCHOOLS MADE UP THE SAMPLE. TEACHERS WERE CONTACTED BY MAIL, SUPPLIED WITH THE MATERIALS AND INSTRUCTIONS, AND DID THE ACTUAL ADMINISTERING OF THE TREATMENTS. INTACT CLASSES WERE RANDOMLY ASSIGNED TO ONE OF THREE CONDITIONS. GROUP 1 RECEIVED PROGRAMED INSTRUCTION PLUS BLOCKED REVIEW. GROUP 2 RECEIVED THE PROGRAM PLUS SPACED REVIEW, AND GROUP 3 THE PROGRAM ALONE. REVIEW WAS PROVIDED BY A SPECIALLY CONSTRUCTED FILMSTRIP. WITH SPACED REVIEW, RELEVANT SECTIONS OF THE FILM WERE SHOWN AFTER EACH SECTION OF THE PROGRAM WAS COVERED. WITH BLOCKED REVIEW THE ENTIRE FILM WAS SHOWN AFTER THE COMPLETION OF THE PROGRAM. THE PROGRAMED TEXTS REQUIRED FROM FIVE TO SEVEN HOURS TO COMPLETE. ALL STUDENTS WERE TESTED FOR AMOUNT LEARNED USING A PAPER-AND-PENCIL TEST CONSTRUCTED BY THE AUTHOR. THEY WERE RE-TESTED FOR RETENTION THIRTY DAYS LATER. THE DATA WERE ANALYZED USING ANALYSIS OF COVARIANCE WITH SCORES FROM A READING TEST AS THE COVARIANT. THE RESULTS DID NOT SUPPORT THE HYPOTHESES. THE ADDITION OF REVIEW TO THE PROGRAMED INSTRUCTION DID NOT LEAD TO SIGNIFICANTLY GREATER LEARNING. BLOCKED REVIEW PROVED BETTER THAN SPACED REVIEW (THE REVERSE OF WHAT WAS HYPOTHESIZED). AND NO DIFFERENCES WERE FOUND BETWEEN TREATMENTS ON THE TEST FOR DELAYED RETENTION. (THE APPENDICES TO THIS STUDY WERE DELETED FROM THIS FINAL REPORT. THEY ARE AVAILABLE FROM THE CORNELL UNIVERSITY LIBRARY.) (RG)

ED013538

A

EM 000 488

FINAL REPORT

Project No. 7-8068

Grant No. OEG-1-7-078068-2993

RECEIVED
ERD

JUN 8 1967

AM 7 8 9 10 11 12 1 2 3 4 5 6 PM

THE RELATIVE EFFECTIVENESS OF SUPPLEMENTING PROGRAMED INSTRUCTION WITH BLOCKED VERSUS SPACED REVIEW

May 1967

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

The Relative Effectiveness of Supplementing Programed Instruction With Blocked Versus Spaced Review

Project No. 7-8068
Grant No. OEG-1-7-078068-2993

James Arnold Scanlon
Frederick K. T. Tom

May 1967

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

Cornell University

Ithaca, New York

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EM 000 488

THE RELATIVE EFFECTIVENESS OF SUPPLEMENTING PROGRAMED INSTRUCTION WITH BLOCKED VERSUS SPACED REVIEW

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University for the Degree of

Doctor of Philosophy

By

James Arnold Scanlon

June 1967

VITA

James Arnold Scanlon was born in Austin, Texas on January 22, 1936. He attended public schools in Texas through the tenth grade then moved to Arkansas where he graduated from Concord High School, Concord, Arkansas, in May 1953.

After enlisting in the United States Marine Corps and serving from July 1954 till May 1957, he entered Southwest Texas State College where he was awarded the B. S. Degree in Vocational Agriculture in May 1961.

He taught vocational agriculture at Monticello, Arkansas, 1961-1962; Lonoke, Arkansas, 1962-1963; and Palestine, Texas, 1963-1964. During this period he completed requirements for the M. Ed. Degree at the University of Arkansas and was awarded this degree January 1963.

He entered Cornell University in September 1964 as a graduate assistant in agricultural education with minors in agricultural engineering and rural sociology. In February 1967 he received an appointment in the Department of Education as a research associate.

On September 13, 1958 he was married to Jo May Stephens of Houston, Texas. Their family includes three children.

ACKNOWLEDGMENTS

This thesis was written under the guidance of the writer's Special Committee composed of the following members:

Professor Frederick K. T. Tom, Department of Education

Professor Fred G. Lechner, Department of Agricultural Engineering

Professor Robert A. Polson, Department of Rural Sociology

The writer wishes to express special gratitude to the chairman of his committee, Dr. Tom, for his untiring efforts and guidance while directing this study.

Acknowledgment with thanks is also expressed to the New York Teachers of Vocational Agriculture who participated in this study and without whose cooperation and efforts this study could not have been conducted.

Sincere appreciation is expressed to my wife, Jo May, for her encouragement and help throughout the period during which this study was conceived and conducted.

Special acknowledgment is given the U. S. Office of Education for Grant No. OEG-1-7-078068-2993, with which the study was financed.

TABLE OF CONTENTS

CHAPTER		Page
I.	INTRODUCTION	1
	Statement of the Problem	2
	Review of Related Research	5
	Theoretical Framework	8
	Importance of the Study	9
	Design of the Experiment	10
II.	METHOD	13
	Selecting the Program	13
	Determining Which Review Procedure to Use	14
	Constructing the Review Filmstrip	14
	Developing the Criterion Test	15
	Educational Objective of the Programed Unit	15
	Selecting the Sample	19
	Obtaining Scores on the Covariable	20
	Assigning Classes to Treatment Groups	20
	Administering the Treatments	21
	Administering the Criterion Test	21
	Analyzing the Data	22

CHAPTER — Page

III. FINDINGS — 24

Learning — 25

Retention — 27

IV. CONCLUSIONS, RECOMMENDATIONS, AND SUMMARY — 29

Conclusions — 29

Recommendations — 30

Summary — 31

BIBLIOGRAPHY — 33

APPENDICES

A. A Reproduction of the Cards from Which the Review Filmstrip was Photographed — 35

B. Criterion Test — 83

C. Instructions for Administering the Nelson-Denny Reading Test, Form A — 90

D. Correspondence: — 93

Initial Contact Letter — 94

Acknowledgment Letter and List of Participants — 96

Letter Regarding Starting Date for Phase II, Administering Treatment — 98

Letter of Instruction for Administering Treatments:

Treatment Group 1 — 99

Treatment Group 2 — 101

Treatment Group 3 — 103

Letter Regarding Date for Administering Retention Test — 104

LIST OF TABLES

TABLE		Page
1.	Percent of Total Emphasis Given to Each Subtopic and Each Objective in the Programed Unit	18
2.	Analysis of Covariance of Scores on Test for Learning	25
3.	Multiple t-test for Differences Between Adjusted Group Means of Learning Test Scores, Adjusted for Variations in Reading Score	26
4.	Analysis of Covariance of Scores on Test for Retention	27

LIST OF ILLUSTRATIONS

FIGURE		Page
I.	Paradigm of the Posttest - Only Control Group Design	11
II.	The Experimental Design	12

CHAPTER I

INTRODUCTION

The last decade has brought about many changes in the field of vocational education. Shifts of emphasis have occurred and new instructional areas have emerged. The advent of Public Law 88-210, the Vocational Education Act of 1963, has furnished the field with the legislative and monetary backing needed for exploring new areas of instruction and for developing new curricula, teaching materials and methods of presentation and dissemination.

The vocational agriculture program, in keeping pace with a rapidly changing society, has become a complex and multi-dimensional educational endeavor involving instruction not only for those who intend to farm, but also for those seeking employment in off-farm agricultural occupations. Expansion has occurred in such a way that in addition to providing instruction in farm production and management, courses have been added in specialty areas such as ornamental horticulture, agricultural mechanization, conservation, forestry, and agricultural business. The development of entry level skill is the prime objective of these new instructional programs just as entry into farming has always been the prime objective of the traditional vocational agriculture program.

This general expansion in the vocational agriculture program has resulted in three shortages: teachers trained in specialty areas;

instructional materials relating to these new areas; and media through which these instructional materials may be presented.

This study is an experiment employing one of these new instructional media. It is designed to determine the relative effectiveness of supplementing programed instruction with blocked versus spaced review.

Statement of the Problem

Since B. F. Skinner's 1954 article, "The Science of Learning and the Art of Teaching," introducing the concept of machine teaching, appeared in Harvard Educational Review, the professional literature in education has been literally flooded with articles and research reports on programed instruction. Programs and program holding devices have also been forthcoming at a rapid rate. Programs, '63[1] lists 352 programs available in ten subject matter areas. This is almost a 300% increase over the 122 programs reported available the previous year.

It is evident at this point that programed instruction is here to stay. Programs are not only becoming more prevalent in the pbulic school, but industry is also employing programed instruction at a rapid rate. For example, Du Pont recently offered, for purchase by the public, ninety programs in industrial subjects.[2] Another factor, the commercialization of programed instruction by publishing companies, and the salesmanship of their representatives, will also contribute to the increased use of programs.

---

1. U. S. Department of Health, Education and Welfare, Programs, '63: A Guide to Programed Instruction (Washington: Government Printing Office, 1963), OE-34015-63, p. vi.

2. "Du Pont Offers Programed Instruction Courses," Du Pont Agricultural News Letter, vol. 33, no. 1 (Spring 1966), p. 3.

Programed instruction has been proven to be an effective method of teaching. Numerous studies have shown that persons learning from programs perform as well, or better, than persons learning from conventional methods of teaching. "It seems clear that teaching machines, like television, books, or a lecturer, can teach. Repeated studies have shown no significant differences in learning by machine as compared to other more conventional modes."[3] As programers develop more skill in their writing, and as improved ways of using programs are discovered, these devices should become even more effective.

Research in the area of programed instruction has been primarily concerned with comparing programed instruction with other teaching methods and with manipulating variables internal to the programs themselves. A definite void exists in that very little research has been conducted on improving the use of existing programs in the classroom. "A common question asked revolves about the role of teaching machines in the total instructional situation. Are they designed to replace the normal instructional pattern, or are they just another aid?"[4] "Our knowledge of how to best utilize them (programs) is meager and imperfect."[5] These are examples of statements found in the literature when one seeks to answer the question of how programs should be used in the classroom.

One recommendation frequently made as to how teachers should incorporate programs into their classroom activities is, "Teachers should

---

3. Egan G. Guba, "Teaching Machines Are Here to Stay," Theory Into Practice, vol. 1, no. 1 (February 1962), p. 6.

4. Desmond L. Cook, "The Automization of Socrates," Theory Into Practice, vol. 1, no. 1 (February 1962), p. 11.

5. Guba, p. 6.

develop, independently of any program, a set of objectives to be attained by students at the end of a particular subject."[6] This statement led the writer to take the following stand with regard to incorporating programed instruction into the overall instructional program in the classroom: If the educational objectives are clearly defined for a particular subject, then the objectives posed for the program (program objectives) may be studied to see which of the educational objectives could be accomplished by administering the program. To the extent the program accomplishes the educational objectives, the program may be used as the sole means of instruction; to the extent the program does not accomplish the educational objectives, the program must be either rejected or supplemented.

Three situations emerge from this stand:

Situation 1--The program alone, accomplishes all of the educational objectives posed for a particular subject.

Situation 2--The program, alone, accomplishes all of its stated objectives, but only some of the educational objectives posed for a particular subject.

Situation 3--The program, alone, only partially accomplishes the objectives for which it was designed.

The third situation is the one with which this experiment is concerned. Whatever the cause of the failure of a program to teach a student, whether the program was not well-written, whether the students were poorly motivated, whether the teacher expected a higher level of achievement than did the program writer, etc., the teacher is somewhat limited in what he

---

6. Richard W. Burns and Mary B. Craik, "Using Programs in the Classroom," Audiovisual Instruction, vol. 10, no. 9 (November 1965), p. 697.

can do to correct the situation. He can either discontinue use of the program, or he can supplement it with additional teaching.[7]

In this situation a common form of supplementation is review, so the teacher asks himself, "Should I or should I not conduct a review In such a case, the question of when the review is to be conducted must be answered. That one should conduct a review and that the review should be spaced rather than blocked will be hypothesized later.

Review of Related Research

A search for related studies was conducted in Education Index, by the H. W. Wilson Company; Research Studies in Education, by Phi Delta Kappa, Inc.; and The Research on Programed Instruction: An Annotated Bibliography, by Wilbur Schramm. Although the body of research done on programed instruction is quite voluminous in size, very little of it can be related to the current experiment because of the nature of the variables treated in this experiment, namely blocked review versus spaced review as a supplement to programed instruction. The search yielded only one study which could be related directly to the current experiment.

Reynolds and others conducted a series of experiments to study the effects of repetition and spaced review on learning and retention from programed instruction. In the introduction of their final report, which is dated December 1964, they make a statement which lends support to the writer in his inability to locate further related studies: "No attempt has been made as yet to determine the effects of repetition and

---

7. It is important for the reader to understand that this study will not deal with teacher-written programs, but with published programs.

spaced review in programed instructional materials."[8]

Spaced review was defined by Reynolds as a sequence of learning conditions in which a second topic is interpolated between successive presentations of the first topic. In his experiments only one topic was reviewed and this review was facilitated by interpolating additional frames of programed instruction, on the topic to be reviewed, between following topics in the program. In other words, only one topic was reviewed and it was reviewed at intervals determined by the length of the following topics.

Although Reynolds' concept of spaced review is slightly different from that held by the writer, there is enough similarity to make his findings relevant to the current experiment. Reynolds found that ". . . spaced review alone rather than repetition and review in combination was found to be the variable facilitating learning and retention."[9] He further stated that ". . . spaced review produced significant increases in learning which persisted and even increased, through a 3-week retention interval."[10]

Reynolds was somewhat reluctant to relate his findings to those of learning theorists and experimental psychologists whose work took place primarily in laboratory settings. He stated, "The consistent and significant effects of spaced review (as found in his experiments) are difficult to relate to previous laboratory research because the manner in which review was distributed in the present studies does not replicate

---

8. James H. Reynolds, et al., Repetition and Spaced Review in Programed Instruction (Ohio: Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, 1964), p. 1.

9. Ibid., p. 26.

10. Ibid., p. 29.

in several respects either the distribution of practice or the retroaction paradigms used in the laboratory."[11]

Studies of the type mentioned by Reynolds involving primarily experimentation in laboratory settings are the ones from which the major portion of the theoretical framework for the current study were derived.

Summaries of the general findings of these studies are presented in "Ten Years of Massed Practice on Distributed Practice" by Underwood,[12] and in a review of the research on retroaction inhibition by Slamecka and Ceraso.[13] The general conclusion regarding review drawn in these two articles was summarized very well by Bohn: "Learning theorists have pointed out and psychologists have demonstrated experimentally that periodic review facilities learning and retention."[14] From this conclusion he hypothesized that "Periodic review during a program helps the students to organize the small bits into a sensible structure." This hypothesis of Bohn's received careful consideration by the writer at the time the two hypotheses for the current study were posed.

The current experiment differed from Reynolds' experiments to approximately the same degree that his differed from previous laboratory

---

11. James H. Reynolds and Robert Glaser, Repetition and Spaced Review in the Learning of Connected Discourse (Washington: U. S. Office of Education, 1963), Project No. 1343, p. 40.

12. B. J. Underwood, "Ten Years of Massed Practice on Distributed Practice," Psychological Review, vol. 68 (1961), pp. 229-247.

13. N. J. Slamecka and J. Ceraso, "Retroactive and Proactive Inhibition of Verbal Learning," Psychological Bulletin, vol. 57 (1960), pp. 449-460.

14. David A. Bohn, "Programed Instruction: A View from the Classroom," Programs, Teachers, and Machines, Alfred de Grazia and David A. Bohn, ed. (New York: Bantam Books, 1964), p. 183.

research in that the way in which review was spaced in the current experiment did not replicate the way in which review was spaced by Reynolds. It further differed in that it compared spaced review with blocked review, it employed a filmstrip in the review procedure, and in the case of the treatment group receiving spaced review, the review immediately followed the programmed topic. It was similar to Reynolds' study in that review was compared with non-review.

Theoretical Framework

The theory for this study incorporated the principles of frequency and whole versus part learning which together gave rise to the following postulates:

1. If there is a constant context within units of the learning task and a disparate context between units of the learning task, learning can be increased by any method that separates the learning task into its separate units.

2. In the case of a lengthy and/or complex learning task, review between units of the task will result in more learning than review only at the end of the entire learning task.

3. The more frequent a learning activity is experienced, the greater is the amount of the resultant learning.

If the assumption is made that there is a constant context within units of the learning task and a disparate context between units of the learning task, then it follows from postulate 1 that a structured review session would be a method of separating the learning task into its separate units. By combining postulates 1 and 2 one could deduce that it would be best to present the review between units of the learning task

rather than to have blocked review at the end of the task. From postulate 3 it follows that review of the learning task would increase the frequency of the learning experience so that one could say that a group having review should do better than a group having no review.

Four general deductions may be drawn from the theory and these are stated as the hypotheses to be tested:

Hypothesis 1: Students who have program texts supplemented with review will learn more than those who have only programed texts, as measured by a test for learning administered immediately after the instructional period.

Hypothesis 2: Students who have programed texts supplemented with spaced review will learn more than those who have programed texts supplemented with blocked review, as measured by a test for learning administered immediately after the instructional period.

Hypothesis 3: Students who have programed texts supplemented with review will retain more than those who have only programed texts, as measured by a test for retention administered after an interval of 30 days.

Hypothesis 4: Students who have programed texts supplemented with spaced review will retain more than those who have programed texts supplemented with blocked review, as measured by a test for retention administered after an interval of 30 days.

Importance of the Study

The findings of this experiment will add to the body of knowledge on the use of review as a supplement to programed instruction. Campbell and Stanley speak of experimentation as: (1) the only means for settling

disputes regarding educational practice, (2) the only way of verifying educational improvements, and (3) the only way of establishing a cumulative tradition in which improvements can be introduced.[15] If this statement is taken as the role of experimentation, the proposed experiment compliments this role in that it will: (1) provide evidence regarding the value of supplementing programs with review, (2) provide data which will demonstrate the superiority of one method of review over another, and (3) add to the body of knowledge in the field in a cumulative tradition by either supporting the theory from which the study is drawn or by supplying evidence contrary to the theory.

The purpose of this study was to test the theory presented on page eight, using programmed instruction. It was desirous to know whether the postulates inherent in the theory are generalizable to programmed learning tasks. The outcomes of this experiment will supply empirical evidence on this.

This stated purpose parallels the goal of educational research workers as expressed by Travers, ". . . to produce a body of knowledge consisting of generalizations about behavior which can be used to predict behavior in educational situations and for planning educational procedures and practices."[16]

Design of the Experiment

The basic model from which the design for this experiment was

---

15. Donald T. Campbell and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research of Teaching," Handbook of Research in Teaching, N. L. Gage, ed., 1963, p. 172.

16. Robert M. W. Travers, An Introduction to Educational Research (New York: Macmillan Co., 1964), p. 6.

constructed is defined by Campbell and Stanley as The Posttest - Only Control Group Design, and is classified, according to them, as a true experimental design.[17] The basic elements of this design are: (1) randomization, (2) treatment, and (3) posttest. This is presented diagrammatically in Figure I, below.

Figure I. Paradigm of the Posttest - Only Control Group Design.

	Treatment Group	Treatment	Posttest
Randomization →	I	Yes	Yes
	Control	None	Yes

In building upon the above design to construct the design for the current experiment, three elements were added: (1) a third treatment group, (2) a measure on a covariable, and (3) a re-test for retention. A diagram of the design for the current experiment is shown in Figure II. The three treatment groups were: (1) programed instruction plus blocked review, (2) programed instruction plus spaced review, and (3) programed instruction only. The dependent variables were student scores on a test for learning and on a re-test for retention given thirty days later. Individual student scores on the Nelson-Denny Reading Test[18] provided a measure of the covariable to be used in the analysis of covariance, the statistical analysis chosen for the experiment.

---

17. Campbell and Stanley, p. 195.

18. M. J. Nelson and E. C. Denny, The Nelson-Denny Reading Test, Revised ed., Form A (Boston: Houghton Mifflin Co., 1960).

Figure II. The Experimental Design.

| Sample | Co-variable | | Treatment Group | Treatment | | |

CHAPTER II

METHOD

The experiment was conducted in a series of rather distinct steps. Step 1 was undertaken immediately after the problem was identified and Step 10 terminated the experiment.

Step 1: Selecting the Program

In selecting an appropriate programed unit for use in the experiment, four criteria were kept in mind: (1) the program must provide instruction on a subject normally taught in vocational agriculture; (2) the program must be of sufficient length so that differences in treatment can be observed with some assurance; (3) the program must be of high quality; and, (4) the subject matter must be difficult enough to be challenging for 9th grade students. With the help of a bibliography of the programed instructional material available for use in secondary schools,[19] Parliamentary Procedure: A Programed Introduction, by John W. Gray and Richard G. Rea, was selected for use in this experiment. This program, which was published in 1963 by Scott, Foresman and Company, meets the above listed criteria. It is in textbook form, linear in format and, requires from five to seven hours to complete.

---

19. The Center for Programed Instruction, Programed Instruction Materials 1964-'65 (Teachers College, Columbia University, 1965).

Step 2: Determining Which Review Procedure to Use

In selecting which review procedure to use in the experiment several criteria were kept in mind, the most important of which was the effect of individual differences among the 29 teachers involved. Since it was impractical, if not impossible, to select teachers with like personalities, a special attempt was made to select a review procedure which was practical, from an economic and technological standpoint, and which would minimize the effect of the teacher during the review. Several review procedures were explored including overhead transparencies, charts, movie films, filmstrips, recordings on both discs and tapes, scripts to be read by the teachers, and various combinations of these. The production expense involved in some of these procedures was prohibitive, and in other cases, the necessary record playing and projection equipment was not available in the schools.

After considering the various alternative procedures, the filmstrip was selected as the best review procedure. Filmstrip projectors were found in all participating schools and the cost of a filmstrip was much less than that of some of the other audio-visuals considered. In essence, the review procedure consisted of the teacher showing his students a series of filmstrip frames on which were contained information the students had earlier learned from the programed textbook. Designated portions of the script contained in the frames were read by the teacher, others by his students.

Step 3: Constructing the Review Filmstrip

In constructing the review filmstrip, care was taken to include only the factual information covered in the programed unit. Each important

concept in the program was identified and filmstrip frames were constructed which would provide a review of each concept. In some cases only a single filmstrip frame was necessary for reviewing a concept; in other cases, two or more frames were deemed necessary. The same ordering of concepts used in the programed unit was used in the filmstrip.

The frames for the review filmstrip were constructed on five-inch by eight-inch index cards. All wording on the cards was typed. A reproduction of the cards from which the filmstrip was photographed is in Appendix A.

The completed cards were sent to George W. Colburn Laboratories, Inc., 164 North Wacker Drive, Chicago 6, Illinois, the firm which actually photographed the filmstrip and which duplicated it in sufficient quantity for use in the experiment.

Step 4: Developing the Criterion Test

A criterion test was developed which would serve two purposes, namely, to measure how much the students learned at the end of the treatment and how much they retained thirty days later. The first step in developing the criterion test was to determine the educational objectives of the program. These objectives were not provided in the programed unit; therefore, it was necessary to determine them, by inference, from the text of the program itself. The objectives are listed below:

Educational Objectives of the Programed Unit

1. When presented with a random list of the eight steps in the order of business, the student should be able to arrange them in their proper order.

2. When presented with an item of business, the student should be able to place it within the proper step in the order of business.

3. When presented with the steps for making and disposing of a motion, the student should be able to arrange them in their correct order.

4. When presented with a parliamentary situation, the student should be able to recognize the correct procedure for handling the situation.

5. The student should be able to recognize a properly worded motion.

6. When presented with a properly worded motion, the student should be able to amend it by each of the three common methods.

7. When presented with a situation in which an amendment to a motion has been amended, the student should know the correct sequence for disposing of the motion and the amendments.

8. The student should be able to recall the four general classifications of motions.

9. The student should be able to classify any given motion.

10. The student should be able to select the proper motion to take care of a given situation which may arise during a meeting.

11. When presented with a random list of all privileged and subsidiary motions, the student should be able to arrange them in their proper order of precedence.

The criterion test used in this experiment was original and unrefined. Although an attempt was made to construct an instrument which would possess validity, there is no assurance, beyond the judgment of the writer, that it measured what it was purported to measure.

In an attempt to establish validity in the criterion test, two types of validity were stressed, content validity and face validity. A test is said to have content validity to the extent that each objective of the unit is emphasized in the test to the same degree that it is emphasized in the unit. A test is said to have face validity to the extent that the test questions appear to measure the degree to which the objectives have been met.

In order to attain content validity, the number of frames in each subtopic of the program was counted and this count was used to determine the weight to be given to each subtopic, or the extent to which each subtopic was to be emphasized in the test. The first subtopic contained 55 frames, the second contained 83, the third 52, and the fourth 404 frames. Each subtopic was given a weight equal to the percent the number of frames on that subtopic was of the total frames of the program. In this manner the four subtopics of the program were given the weights nine, fourteen, nine, and sixty-eight respectively. Similarly, each subtopic was divided among its objectives accordingly. The table below shows the emphasis given to each objective as determined by the above procedure.

Using the table below as a guide, the test question developed to evaluate a particular objective was given the same emphasis as the objective. For example, the test question for objective three was given a weight of four percent of the total test score of one hundred.

Face validity was striven for by selecting the type of test question which, in the judgment of the writer would best evaluate a particular objective. Since the objectives were stated rather explicitly in terms of the terminal behavior it was expected the students should display when

Table 1. Percent of Total Emphasis Given to Each Subtopic and Each Objective in the Programed Unit.

Subtopic Number	Percent of Total Emphasis	Objective Number	Percent of Total Emphasis
1	9%	1	3%
(55 frames)		2	6%
2	14%	3	4%
		4	6%
(83 frames)		5	4%
3	9%	6	3%
(52 frames)		7	6%
4	68%	8	4%
		9	24%
		10	30%
(404 frames)		11	10%

they finished the program, test questions were constructed which called upon the student to display this terminal behavior. For example, objective number one was: When presented with a random list of the eight steps in the order of business, the student should be able to arrange them in their proper order. The test question constructed to evaluate the degree to which this objective was accomplished was test question number one which read: Indicate the correct order of the above steps by filling in the blanks below with the appropriate letter. Above the question were listed in random order, the eight steps in the order of business, each preceeded by a small letter; below the question the students were provided with eight blanks labeled Step 1 through Step 8. The placing of the correct small letters in the proper blanks was the

terminal behavior called for, within the limitations of a paper-pencil test.

An inherent element in this experiment is the basic assumption that the procedures used by the writer in devising the evaluative instrument (paper-pencil test) resulted in a valid instrument. A copy of the test is in Appendix B.

Step 5: Selecting the Sample

Ninth grade students enrolled in vocational agriculture in New York were selected as the subjects of this experiment. Parliamentary procedure is usually taught at the ninth grade level to students of vocational agriculture as part of the leadership training activities of the Future Farmers of America Organization.

All vocational agriculture departments operated by boards of education in New York were given a different number. Then, enough schools were selected at random, using a table of random numbers, until approximately 350 students were secured as subjects. Some of the schools selected were unable to participate. The final sample involved 29 teachers in 28 schools.

Some of the teachers reported that whereas their first year agriculture class was composed of predominantly ninth graders, there were some tenth, eleventh and twelvth grade students included also. While all students within the classes selected were allowed to participate, only data from ninth and tenth grade students were used in the analysis. The total number of students from whom complete data were collected was 279, of which 228 were ninth graders and 51 were tenth graders. The use of these few tenth grade students was justified by using reading ability

as a statistical control in the analysis.

Step 6: Obtaining Scores on the Covariable

All students were tested for reading ability using Form A of the Nelson-Denny Reading Test. This test yields a score based on the individual's vocabulary and reading comprehension ability. A third part of the test, reading rate, was considered of no importance because the students were to be allowed to proceed through the program at their own pace. This test was administered in the standard manner by the cooperating teachers well in advance of the actual starting date for the treatment. A copy of the instructions for administering the reading test is in Appendix C.

Step 7: Assigning Classes to Treatment Groups

The twenty-nine classes were ranked from one to twenty-nine on mean reading score. Since class size varied widely, the treatment groups were adjusted for number as the classes were assigned to them. For example, the five highest classes on mean reading score contained seven, seventeen, five, ten, and twelve students, respectively. For the purpose of assignment to treatment groups, the class containing five students was combined with the class of twelve, and the class of seven students was combined with the class of ten. This resulted in three groups of seventeen students each. The groups were then assigned at random to one of the three treatment groups. Such a procedure was followed until all classes were assigned to a treatment.

This method of assignment to treatment groups met the requirements for random assignment and also insured that the three groups were composed of students of approximately equal reading ability.

Step 8: Administering the Treatments

The teachers were contacted by letter relative to the date for starting the treatment. Four starting dates were proposed and each teacher was asked to indicate his choice on a return postcard. The four dates were January 9th, 16th, 23rd, and 30th, of 1967. This selection allowed the teacher to plan ahead for the purposes of both finishing any pending instructional units and also to select a time during which there would be the least number of interruptions.

In administering the treatments, the group having spaced review completed one segment of the program, then had a review of that segment by viewing the corresponding segment of the filmstrip. This pattern was continued until the instructional period was over. The group having blocked review completed the entire program, then viewed the entire filmstrip. The group having program only did not view the filmstrip.

The instructional materials were mailed to the teachers along with explicit instructions for administering the appropriate treatment to their students. Copies of the letters containing those instructions are in Appendix D. All instructional materials were donated to the participating schools upon termination of the experiment.

Step 9: Administering the Criterion Test

The test for learning was administered by the teacher during the first whole class period after the treatment was completed. The re-test for retention was administered thirty days after the first test. In cases where the thirtieth day fell on a weekend, the retention test was given on the preceeding Friday or the following Monday. The test time was standardized at forty minutes.

Step 10: Analyzing the Data

Scores on the three variables of interest, namely, reading ability, amount of learning, and amount of retention, were punched on IBM cards so that the data analysis could be facilitated by using a Control Data 1604 Computer housed in the Cornell University Computing Center in Ithaca. A program called "UNIVAR" was available at the Center which was designed to handle a one-way analysis of covariance and a multiple t-test of means, as adjusted by covariance. The mathematics of these procedures may be found in Federer.[20]

Two one-way analyses of covariance were run, the first on learning test scores adjusted for variations in reading score, and the second on retention test scores adjusted for variations in reading score. The covariance analyses provided the writer with F values which were compared with tabled values of F to determine significance. The F value is computed by dividing the mean square (or the variance) between groups by the mean square (or the variance) within groups. Since it was hypothesized that there would be greater variation between groups than within groups, the F value should be greater than one. If the F value is as large or larger than the tabled value of F for a predetermined level of probability, then a significant difference exists between the groups. This difference would be attributable to treatment effects.

A multiple t-test of adjusted mean test scores of the three treatment groups was run simultaneously with the analysis of covariance. The multiple t-test of adjusted means yielded an LSD (Least Significant Difference) value which takes the place of the tabled values of t used in the

---

20. Walter T. Federer, Experimental Design (New York: The Macmillan Company, 1955), pp. 20-21 and p. 485.

more common t-test. If the difference between any two adjusted means is equal to or larger than the LSD value computed for that pair of means, then a significant difference exists between the adjusted means.

Throughout the experiment the .05 level of probability was used to determine if significant differences existed. When it is said that differences are significant at the .05 level of probability, that means the differences could have happened by chance no more than five times in 100.

CHAPTER III

FINDINGS

The primary objective of this experiment was to determine the relative effectiveness of supplementing programmed instruction with blocked versus spaced review. Four deductions were drawn from the theoretical framework and posed as the hypotheses around which the experiment was designed. These are reproduced here for the purpose of convenient reference.

Hypothesis 1: Students who have program texts supplemented with review will learn more than those who have only programmed texts, as measured by a test for learning administered immediately after the instructional period.

Hypothesis 2: Students who have programmed texts supplemented with spaced review will learn more than those who have programmed texts supplemented with blocked review, as measured by a test for learning administered immediately after the instructional period.

Hypothesis 3: Students who have programmed texts supplemented with review will retain more than those who have only programmed texts, as measured by a test for retention administered after an interval of 30 days.

Hypothesis 4: Students who have programmed texts supplemented with spaced review will retain more than those who have programmed texts

supplemented with blocked review, as measured by a test for retention administered after an interval of 30 days.

Since the two variables of interest were learning and retention, it was convenient to structure this chapter in terms of these two variables.

Learning: The first analysis done was an analysis of covariance on learning test scores adjusted for variations in reading score. This analysis resulted in a significant F value at the .05 level of probability, meaning that there was greater variation between groups than within groups. This difference was attributed to treatment effects. The results of this analysis are presented below, in Table 2.

Table 2. Analysis of Covariance of Scores on Test for Learning.

Sources of Variation	Degrees of Freedom	Sums of Squares	Mean Squares
Between Groups	2	1026.50	513.25
Within Groups	275	40416.00	146.97
Total	277	41442.50	

$F.05 = 3.04$ (with 2 and 275 degrees of freedom)

$F = \frac{513.25}{146.97} = 3.49$

Although the above analysis of covariance yielded a significant F value, it did not show between which of the three groups was the

significant difference attributable. For this, a multiple t-test was run between all possible pairs of the three adjusted group means in order to locate the one or more pairs of means which were significantly different from each other. Table 3 shows the results of this multiple t-test.

Table 3. Multiple t-test for Differences Between Adjusted Group Means of Learning Test Scores, Adjusted for Variations in Reading Score.

Group	Adjusted Means	Differences Between Means	Least Significant Difference*
1 (Program plus blocked review, n = 99)	52.51	$\bar{X}_1-\bar{X}_2=4.66$	4.04
2 (Program plus spaced review, n = 88)	47.85	$\bar{X}_3-\bar{X}_2=1.96$	4.20
3 (Program only, n = 92)	49.81	$\bar{X}_1-\bar{X}_3=2.70$	4.05

*At the .05 level of probability

From Table 3 it can be seen that there was no significant difference between the means of Group 1 and Group 3 and between those of Group 2 and Group 3, as revealed by the fact that the differences between the adjusted means in question were less than the values required at the .05 level of probability. A significant difference did exist, however, between Group 1 and Group 2 with the adjusted mean for Group 1, 52.51, being significantly higher than that for Group 2, 47.85. This is indicated by the difference of 4.66 being greater than the Least Significant Difference of 4.04.

On the basis of this first analysis Hypothesis 1, stating that the groups having review will learn more than the group having only the program, was not accepted because neither of the groups which had the program supplemented with review did better than the group which had only the program. Hypothesis 2, stating that the group having spaced review will learn more than the group having blocked review, was rejected because the finding was opposite to that hypothesized; that is, the group having blocked review did significantly better than the group having spaced review.

Retention: A second analysis of covariance was run on retention test scores adjusted for variations in reading score. This analysis yielded a non-significant F value at the .05 level indicating no significant difference in performance between any two of the three groups. This analysis is presented in Table 4. The adjusted mean scores of the three groups on the retention test were: Group 1 - 49.1, Group 2 - 47.3, and Group 3 - 49.9.

Table 4. Analysis of Covariance of Scores on Test for Retention.

Sources of Variation	Degrees of Freedom	Sums of Squares	Mean Squares
Between Groups	2	308.22	154.11
Within Groups	275	39564.00	143.87
Total	277	39872.22	

P.05 = 3.04 (with 2 and 275 degrees of freedom)

$$F = \frac{154.11}{143.87} = 1.07$$

The fact that the F value in this analysis was just more than 1 means that there was almost as much variation within groups as between groups, and any treatment effect which was present during the test for learning was no longer present 30 days later, during the test for retention. On the basis of this second analysis Hypothesis 3, stating that students having review will retain more than those having program only, and Hypothesis 4, stating that students having spaced review will retain more than those having blocked review, were not accepted.

CHAPTER IV

CONCLUSIONS, RECOMMENDATIONS, AND SUMMARY

The findings of this study reported in the preceding chapter were discussed in terms of learning, retention, and the four hypotheses posed. It remains for the writer to synthesize from these findings concise and logical conclusions. It should be remembered, however, that the conclusions are drawn only within the realm of the findings of this experiment and should be interpreted only within that framework. Included also in this chapter are recommendations and a summary.

Conclusions

Conclusion 1: Supplementing a program with either spaced or blocked review has no beneficial effect in either learning or retention, over administering the program alone.

Conclusion 2: Supplementing a program with blocked review will result in more learning than supplementing it with spaced review although it does not produce a beneficial effect in retention.

Since the writer is unaware of any other study which this experiment replicates, the conclusions cannot be compared with others. A comparison can be made, however, with the conclusion drawn by Reynolds in his series of experiments with spaced review, as reported in Chapter I. Reynolds concluded that "Spacing of review sequences between interpolated

learning material facilitated retention."[21] The current experiment failed to demonstrate this; however, it should be remembered that this experiment employed both a different definition and means of review.

Recommendations

Two types of recommendations were considered appropriate for inclusion in the final chapter of this thesis. These are recommendations for additional research and recommendations for supplementing programed instruction with review. Because of the importance of the problem which resulted in the construction of the current experiment and because of the dearth of studies dealing with spaced and blocked review as a supplement to programed instruction, the writer recommends that additional experiments be conducted in this area. This research may take the form of replications of the current experiment with variations of such factors as subject matter of the program, method of review, grade level of the sample, the covariable, or others, according to the judgment of the researchers.

The writer's recommendation regarding the use of review as a supplement to programed instruction was influenced by the soundness of the theoretical framework used on the basis for the current experiment and by the lack of other evidence supporting the negative findings of the current study. On these bases the writer recommends that in a situation where the teacher finds it necessary to supplement a program with review, he employ spaced review.

---

21. Reynolds and Glaser, p. 39.

Summary

This study involved an experiment, the purpose of which was to determine the relative effectiveness of supplementing programed instruction with blocked versus spaced review. It was hypothesized that: (1) students who have program texts supplemented with review will learn more than those who have only programed texts, as measured by a test for learning administered immediately after the instructional period; (2) students who have programed texts supplemented with spaced review will learn more than those who have programed texts supplemented with blocked review, as measured by a test for learning administered immediately after the instructional period; (3) students who have programed texts supplemented with review will retain more than those who have only programed texts, as measured by a test for retention administered after an interval of 30 days; and (4) students who have programed texts supplemented with spaced review will retain more than those who have programed texts supplemented with blocked review, as measured by a test for retention administered after an interval of 30 days.

Students enrolled in first-year vocational agriculture in twenty-seven New York Schools, selected at random, made up the sample. These students were first tested for reading ability, then intact classes were randomly assigned to one of three treatment groups using a procedure which insured groups of approximately equal mean reading ability. These treatment groups were: Group 1 - program plus blocked review; Group 2 - program plus spaced review; and Group 3 - program only. Review was provided by a filmstrip made especially for use in the experiment. The group having spaced review completed one segment of the program, then had a review of that segment by viewing the corresponding segment of the filmstrip. This

pattern was continued until the instructional period was over. The group having blocked review completed the entire program, then viewed the entire review filmstrip. The group having program only did not view the filmstrip.

At the end of the instructional period, all students were tested for learning using an objective type paper-pencil test constructed by the writer. They were re-tested for retention after a thirty-day interval. Complete sets of data, which were made up of a reading test score, a learning test score, and a retention test score, were collected on 279 students of which 228 were ninth graders and 51 were tenth graders.

The data were analyzed by two one-way analyses of covariance, the first being run on learning test scores adjusted for variations in reading score, and the second on retention test scores adjusted for variations in reading score. Neither spaced review nor blocked review resulted in significantly better performance than program only as measured by the learning test or by the retention test. Blocked review was significantly better than spaced review as measured by the learning test, but not by the retention test.

BIBLIOGRAPHY

A. BOOKS

Campbell, Donald T. and Julian C. Stanley. "Experimental and Quasi-Experimental Designs for Research on Teaching" in N. L. Gage, ed., Handbook of Research on Teaching. Chicago: Rand McNally and Company, 1963. 1,217 pp.

Federer, Walter T. Experimental Design: Theory and Application. New York: The Macmillan Company, 1955. 544 pp.

Gray, John W. and Richard G. Rea. Parliamentary Procedure: A Programed Introduction. Chicago: Scott, Foresman and Co., 1963. 117 pp.

Programed Instructional Materials 1964-'65. New York: The Center for Programed Instruction. Teachers College, Columbia University, 1965. 164 pp.

Sohn, David A. "Programed Instruction: A View from the Classroom" in Alfred de Grazia and David A. Sohn, ed., Programs, Teachers, and Machines. New York: Bantam Books, 1964. 309 pp.

Travers, Robert M. W. An Introduction to Educational Research, 2nd ed. New York: The Macmillan Company, 1964. 581 pp.

B. BULLETINS AND PAMPHLETS

Reynolds, James H., et al. Repetition and Spaced Review in Programed Instruction. Ohio: Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, 1964. Bulletin No. AMRL-TR-64-128. 28 pp.

Reynolds, James H. and Robert Glaser. Repetition and Spaced Review in the Learning of Connected Discourse. Washington: U. S. Office of Education, Cooperative Research Program, December, 1963. Project No. 1343. 45 pp.

U. S. Department of Health, Education and Welfare. Program, '63: A Guide to Programed Instructional Materials. Washington: Government Printing Office, 1963. Bulletin No. OE-34015-63. 785 pp.

C. PERIODICALS

Burns, Richard W. and Mary B. Craik. "Using Programs in the Classroom." Audiovisual Instruction: vol. 10, no. 9 (November 1965), pp. 697-699.

Cook, Desmond L. "The Automization of Socrates." Theory Into Practice: vol. 1, no. 1 (February 1962), pp. 9-19.

"Du Pont Offers Programed Instruction Courses." Du Pont Agricultural News Letter: vol. 33, no. 1 (Spring 1966), p. 3.

Guba, Egan G. "Teaching Machines Are Here to Stay." Theory Into Practice: vol. 1, no. 1 (February 1962), pp. 1-6.

Skinner, B. F. "The Science of Learning and the Art of Teaching." Harvard Educational Review: vol. 24 (Spring 1954), pp. 86-97.

Slamecka, N. J. and J. Cerasco. "Retroactive and Proactive Inhibition of Verbal Learning." Psychological Bulletin: vol. 57 (1960), pp. 49-60.

Underwood, J. B. "Ten Years of Massed Practice on Distributed Practice." Psychological Review: vol. 68 (1961), pp. 229-247.

D. OTHER SOURCES

Nelson, M. J. and E. C. Denny. The Nelson-Denny Reading Test, Form A, Revised ed. Boston: Houghton Mifflin Co., 1960. 9 pp.

APPENDIX

(The appendices of this thesis have been deleted for the purposes of this final report. If it is desired to view the appendix material, the reader is referred to the complete thesis available in the Cornell University Library.)